

Notes and discussion

Things to remember when transcribing speech

David Crystal
University of Reading

Until the day comes when this journal is available in an audio or video format, we shall have to rely on transcriptions to represent the spoken language of the children, and of those who talk to them. There are many kinds of transcription, but they all have the same aim of getting the speech down on paper sufficiently accurately that we no longer need to listen to the tape, to discover what went on. When we read a good transcription, it should be possible to 'hear' the speech of the participants.

Transcriptions can be made at many levels of detail. Some are highly detailed, with the transcriber aiming to write down every audible sound; others are much less detailed, and capture only the most general features of what went on. Which level of transcription satisfies you will depend on your purpose: if you are working on subtle features of a child's pronunciation, you will need to make a precise phonetic transcription; on the other hand, if you are working solely on vocabulary, it may not be necessary to have any phonetic information present in the transcription.

The least specialized transcription is usually referred to as **ORTHOGRAPHIC**. Here, speech is represented using normal spelling and

punctuation, in such a manner as follows:

T What can you see in the picture, John?

P A big car. *P points to car*

T Yes, that's right.

It's a car.

Note that any transcription will introduce certain conventions of layout. In the above illustration, each speaker is given a new line; the names are abbreviated on the left, and an anonymous designation used (T stands for teacher/therapist; P for patient/pupil). Any relevant contextual information, required to make sense of what is going on, is added on the right.

This is a fairly straightforward method, widely used in work on language acquisition and handicap, but it has several limitations. In particular, it tells us nothing about the silences used by the participants, or about the intonation (the speech melody) and rhythm with which they were speaking. Did P pause before he replied, in the above example? Did he stress the word *big*, or *car*, or both? Which word did T stress – *you* or *see* or *picture*? Sometimes, the lack of these indications (what are usually referred to as PROSODIC FEATURES) can make an utterance ambiguous. The word *yes*, written like that, does not tell us whether it was pronounced with a strongly affirmative force, or with a doubtful, dubious force. Nor does the orthographic transcription *he washed and brushed his hair* tell us whether one entity was washed (the hair) or two (i.e. he washed (his face) and (then) brushed his hair). To make these matters clear, we would have to use a PROSODIC transcription.

A prosodic transcription uses the same general layout conventions as the above, but replaces punctuation by symbols to indicate the pauses, rhythm, intonation and other features of speech. Many different methods have been suggested. The one I usually use myself involves the following conventions:

(i) Pauses are marked by a system of dots and dashes: a dot is the shortest possible pause; a dash is a longer pause, equivalent to a unit of the speaker's rhythm (usually lasting about a second); and longer pauses are signalled by sequences of dashes.

(ii) The main rhythm units of speech are separated by a vertical line.

(iii) The louder words in the speech (the stressed words) are marked with a small vertical line placed above and before the word (or before the strongest syllable, if the word has more than one syllable).

(iv) The most prominent word of all within the rhythm unit is written in capital letters.

If we turn the above dialogue into these conventions, it looks like this:

T 'what can you 'see in the PICTURE 'John/

which has the rhythm 'tum-te-te tum-te-te TUM-te tum'. John then replies:

P a 'big CAR/

which has the rhythm 'te-tum TUM'.

These rhythm units are often called *intonation units*, or *tone units*, because they have a distinctive melodic pattern, as well as a distinctive rhythm. The prominence of the word in capitals is in fact usually due to the way the pitch of the voice moves within it – falling, rising, staying level, or moving up and down in more complex ways. These movements are often written in, as follows:

T 'what can you 'see in the PICTURE 'John/

The accent here indicates that the voice pitch is rising, as in a query. The choice between uttering a sentence with a rising pitch and with a falling pitch can dramatically alter the meaning. Compare:

he's 'coming TOMÒRROW/vs he's 'coming TOMÓRROW/.

The sentence with the falling pitch carries the nuance 'I'm telling you'; the sentence with the rising pitch means 'I'm asking you' – or perhaps 'I'm surprised at what you've told me'. Once we add accents to words in this way, it is no longer so necessary to write the most prominent word in capital letters: the accent mark by itself can signal which word this is, as in:

he's 'coming tomòrrow/.

The full prosodic transcription of the above dialogue, as it might have been said, is therefore as follows:

T 'what can you 'see in the picture 'John/

P --- a 'big -- càr/

P points to car

T yès/

that's ríght/ -

it's a càr/

Here, the dashes tell us that P may have a word-finding problem, or at least have some difficulty in processing complex phrases.

It should also be noted that in this example of a prosodic transcription, capital letters are not used at the beginnings of sentences – except for proper names and the word *I*. There are other systems which do retain the capital letters, and which use different conventions for marking the pitch contrasts – such as putting the accent mark BEFORE the word it applies to (as in '*picture*'), or using numbers in place of accents (as in ³*picture*¹ – which would mean that the melody is at pitch level 3 at the beginning of the word, and has moved to pitch level 1 at the end). This numerical system is quite popular in American transcriptions. It is always important

to indicate whose system of prosodic transcription you are using, in this respect, so that the symbols can be interpreted: 3-1 might mean that the pitch is falling (if 3 is high and 1 is low) or rising (if 3 is low and 1 is high).

Both orthographic and prosodic transcriptions are useful as a basis for general work on vocabulary, grammar and pragmatics; but neither suffices for work on pronunciation. Here, a specialized phonetic transcription is required, the most well-known one using the conventions of the International Phonetic Alphabet (IPA). This alphabet uses a different symbol for every distinct sound. So, the two sounds which are spelled identically by English *th* (as in *thin* and *this*) are differentiated in the IPA by two symbols, [θ] and [ð] – the square brackets being used to indicate that phonetic symbols are involved.

The IPA contains a large number of symbols, for it is designed to cope with the sounds in all the world's languages. But when we are transcribing a single language, such as English, it is not necessary to use all of the symbols, as long as we have enough to cover all the important contrasts of sound (the PHONEMES) in the various accents of English. Depending on the accent, it usually takes around 40 symbols to capture all the distinctions of sound that are written using the 26 letters of the ordinary alphabet. The main difference between the two alphabets lies in the vowels: as there are only five vowel letters to cope with around 20 vowel sounds, it is here that the phoneticians have been most inventive. So in the PHONEMIC TRANSCRIPTION of the above dialogue (that is, a transcription which writes down only the most important units of sound), it is the vowel symbols which will seem strangest, when this kind of approach is encountered for the first time.

T 'wɒt kæn ju: 'si: ɪn ðə pɪktʃə 'dʒɒn/

P ə 'bɪg kɑ: /

T jès/

ðæts rài/

ɪts ə kɑ: /

Two points should be appreciated about this kind of transcription. First of all, lots of tiny phonetic details have been left out, in order to focus on the essential features of the pronunciation. The transcription does not indicate the many small movements of the vocal organs which are taking place – movements which might make quite a difference to the sound of P's speech, if he has a pronunciation problem. For example, if P had poor control of his soft palate, and his vowel sound came out in a very nasal way, we could indicate this by using a special phonetic symbol, as follows:

[kã:]. The transcription of children with severe pronunciation problems often ends up looking very complex, accordingly.

Secondly, it must not be forgotten that the IPA is only ONE way of transcribing speech; there are other sets of symbols in existence. Moreover, even within the IPA framework, there is room for difference of opinion about which symbols are best to use. A sound is a complex thing, produced by the combination of many different movements of the vocal organs. Phoneticians may therefore disagree about which of these movements are the most important, and transcribe the 'same' sound in slightly different ways. For example, the vowel sounds of *seat* and *sit* are plainly different, in southern British English. But there are at least two features which make the difference: the vowel of *seat* is longer than that of *sit*, and it is also produced higher up in the mouth. If you think the essential difference between the two words is the vowel length, you can use a length symbol (:) alone to differentiate them, thus: /si:t/ vs /sit/ – and this is what some phoneticians do. Others, however, think that the difference in tongue height is just as important as the difference in length, and they symbolize both these facts in their choice of transcription, thus: /si:t/ vs /sit/. The same pronunciation of the word *sit* thus ends up with competing transcriptions. Once again, you need to know whose system you are dealing with, when you read a phonemic transcription.

It should be plain, from this account, that there are many levels and kinds of speech transcription, ranging from a rough jotting down of the essential words to a meticulous description of everything that can be heard. Which you choose to do will depend partly on your purpose, partly on your training. But choose you must, if you wish to have any kind of objective record of children's language.

Further reading

- Crystal, D. 1982: *Profiling linguistic disability*. London: Edward Arnold.
 Gimson, A. C. 1980: *An introduction to the pronunciation of English* (3rd edn). London: Edward Arnold.
 Grunwell, P. *et al.* 1980: The phonetic representation of disordered speech. *British Journal of Disorders of Communication* **15**, 215–220.
 International Phonetics Association 1979: IPA Chart. University College London.
 O'Connor, J. D. 1973: *Phonetics*. Harmondsworth: Penguin.

(Received 1 June 1984)